The Monge–Ampère equation: various forms and numerical solution

V. Zheligovsky[a,b], O. Podvigina[a,b], U. Frisch[b]

[a]*International Institute of Earthquake Prediction Theory and Mathematical Geophysics,*
84/32 Profsoyuznaya St, 117997 Moscow, Russian Federation
[b]*UNS, CNRS, Laboratoire Cassiopée, Observatoire de la Côte d'Azur*
BP 4229, 06304 Nice Cedex 4, France

Abstract

We present three novel forms of the Monge–Ampère equation, which is used, e.g., in image processing and in reconstruction of mass transportation in the primordial Universe. The central role in this paper is played by our Fourier integral form, for which we establish positivity and sharp bound properties of the kernels. This is the basis for the development of a new method for solving numerically the space-periodic Monge–Ampère problem in an odd-dimensional space. Convergence is illustrated for a test problem of cosmological type, in which a Gaussian distribution of matter is assumed in each localised object, and the right-hand side of the Monge–Ampère equation is a sum of such distributions.

Key words: Monge–Ampère equation, numerical solution, iterative methods
PACS: 02.60.Cb, 02.70.-c

1. Introduction

The Monge–Ampère equation (MAE)

$$\det \|u_{x_i x_j}\| = f(\mathbf{x}) \tag{1}$$

is encountered in many areas of numerical analysis and physics, ranging from image processing [1–3] to cosmology [4–6]. Here the subscripts x_i denote derivatives in the respective spatial variables; $\|u_{x_i,x_j}\|$ is the Hessian of u, i.e., the matrix comprised of second derivatives of u. Existence and regularity of its solutions was considered in [7–9]. Various strategies were proposed for its numerical solution. A provably convergent method for solving the Dirichlet problem for the two-dimensional MAE in bounded convex domains was presented in [10]. It is directly linked to the geometric interpretation of the equation that had given an opportunity to demonstrate existence of weak solutions [13]. However, the actual numerical examples in [10] have just up to 25 grid points, which is clearly insufficient to establish the practical feasibility of the method.[1] The MAE can be recast as a minimisation problem; application of algorithms for saddle-point optimisation to a two-dimensional MAE was considered in [14, 15], and least-squares minimisation was advocated in [16]. Efficient methods for solution of the discrete optimal transportation problem were developed for application to cosmological problems [4–6]. It may be perceived that discrete methods correspond well to the physics of mass transportation in the Universe — after all, galaxies, as observed by astronomers, are clearly well-localised, discrete objects! This argument, however, becomes less persuading, when one recalls that visible matter constitutes only a very small fraction of different, "dark" kinds of matter, whose density distribution is supposed to be continuous (as opposed to a nearly discrete distribution in clusters of the visible matter).

---

*Corresponding author

Email addresses: vlad@mitp.ru (V. Zheligovsky), olgap@mitp.ru (O. Podvigina), uriel@oca.eu (U. Frisch)

[1]More recently, a similar algorithm was used [11] to design a beam–shaping reflector; computations on a 55×55 grid required about 15 minutes of a Pentium 4 processor. This problem of geometric optics can be recast as the Monge–Kantorovich mass transfer problem, reducing to a PDE of the MAE type, see [12].

In another group of methods the MAE is treated as a generic nonlinear partial differential equation. Application of Galerkin's and finite element methods to a regularised MAE was considered in [17, 18]. A pseudospectral Newton's algorithm was reported to perform well for a MAE in R^2 with a smooth r.h.s. [19]. However, when we implemented this algorithm for a three-dimensional MAE, we found that it failed to converge, unless the r.h.s. was a smooth slowly varying function. This is consistent with the observation [10] that Newton's method, applied to equations equivalent to the MAE, is useful to improve an approximate solution only if the approximation is accurate enough, and easily fails otherwise. Two methods for numerical solution of a Dirichlet problem for an elliptic MAE in a two-dimensional convex region Ω are examined in [20]. The first one employs a finite-difference discretisation of the equation; it is advocated for application to the MAE with a possibly singular solution. The second one is an iterative method for the MAE in the form of a fixed point problem

$$ u = \nabla^{-2} \sqrt{u_{x_1 x_1}^2 + u_{x_2 x_2}^2 + 2 u_{x_1 x_2}^2 + 2f} $$

(here ∇^{-2} denotes the inverse Laplacian); it is claimed to perform better, when the solution is regular (i.e., belongs to the Sobolev space $W_2^2(\Omega)$).

A so-called "inexact" iterative Newton–Krylov solver with preconditioning was applied for two- [21] and three-dimensional [22] grid generation with the properties of equidistribution and minimum distortion. The nature of the problem solved in [21, 22] required only a modest accuracy of solutions, discrepancies the order of $10^{-3} - 10^{-4}$ being acceptable. Examples of computations of such moderately accurate solutions illustrating the scalability of the algorithm with respect to the employed spatial resolution were provided *ibid.*

In Sections 2-4 we derive three alternative forms of (1): the "second-order divergence" form, the Fourier integral form and the "convolution" form, which, to the best of our knowledge, were never presented in the literature before. The second form suggests two related methods for computation of space-periodic solutions to (1). The methods and results of their test applications are presented in Section 6; the Monge-Ampère problem of the cosmological type, which we use to test our algorithms, is presented in Section 5, after a short statement of the cosmological reconstruction problem. In the Concluding remarks we briefly discuss some open questions.

We note that the MAE's of various kinds are considered in the literature. For the MAE arising in the mass transportation problem, typically the r.h.s. of (1) depends on the gradient of the unknown function (this form of the equation can be found in the original memoir by Ampère [24]). In the theory of PDE's, usually one considers the MAE with a known r.h.s. (see [9, 23]). The MAE of this kind arises in the reconstruction of the early Universe (see discussion at the beginning of Section 5). Since this cosmological problem is the main original motivation for our work, we restrict our discussion to the case of the "mathematician's" MAE. However, a straightforward reformulation of our algorithms makes them applicable for the case of the r.h.s. depending on the unknown function.

2. The "second-order divergence" form of the MAE

The divergence form of the MAE (1), in which the l.h.s. of the equation is represented as a sum of the *first* derivatives of certain quantities, is well known (see, e.g., [17]). In this section we derive a representation of the l.h.s. of the MAE as a sum of *second* derivatives, which we call the second-order divergence form of the MAE.

Consider a Fourier integral solution

$$ u = \int_{R^N} \widetilde{u}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}\cdot\mathbf{x}} \, d\boldsymbol{\omega} $$

to the MAE in R^N. Substituting the integral into (1) and using the identity for $N \times N$ matrices

$$ \det \|a_{ij}\| = \frac{1}{N!} \sum_{i_1,\ldots,i_N,j_1,\ldots,j_N} \varepsilon_{i_1\ldots i_N} \varepsilon_{j_1\ldots j_N} \prod_{n=1}^{N} a_{i_n j_n} $$

where each of the indices $i_1, ..., i_N, j_1, ..., j_N$ takes the values $1, ..., N$ and $\varepsilon_{p_1...p_N}$ denotes the unit antisymmetric tensor of rank N, we find

$$\det \|u_{x_i x_j}\| = \frac{(-1)^N}{N!} \sum_{i_1,...,i_N,j_1,...,j_N} \varepsilon_{i_1...i_N} \varepsilon_{j_1...j_N} \prod_{n=1}^{N} \int_{R^N} \widetilde{u}(\boldsymbol{\omega}) \omega_{i_n} \omega_{j_n} e^{i\boldsymbol{\omega}\cdot\mathbf{x}} \, d\boldsymbol{\omega}$$

$$= \frac{(-1)^N}{N!} \int_{R^N} ... \int_{R^N} \left(\sum_{i_1,...,i_N} \varepsilon_{i_1...i_N} \prod_{n=1}^{N} \omega_{i_n}^n \right) \left(\sum_{j_1,...,j_N} \varepsilon_{j_1...j_N} \prod_{n=1}^{N} \omega_{j_n}^n \right) \left(\prod_{n=1}^{N} \widetilde{u}(\boldsymbol{\omega}^n) \right)$$

$$\times \exp\left(i \sum_{n=1}^{N} \boldsymbol{\omega}^n \cdot \mathbf{x} \right) d\boldsymbol{\omega}^1 ... d\boldsymbol{\omega}^N$$

$$= \frac{(-1)^N}{N!} \int_{R^N} ... \int_{R^N} \det{}^2 \left\| \boldsymbol{\omega}^1, ..., \boldsymbol{\omega}^{N-1}, \boldsymbol{\omega} - \sum_{n=1}^{N-1} \boldsymbol{\omega}^n \right\|$$

$$\times \left(\prod_{n=1}^{N-1} \widetilde{u}(\boldsymbol{\omega}^n) \right) \widetilde{u}\left(\boldsymbol{\omega} - \sum_{n=1}^{N-1} \boldsymbol{\omega}^n \right) e^{i\boldsymbol{\omega}\cdot\mathbf{x}} \, d\boldsymbol{\omega}^1 ... d\boldsymbol{\omega}^{N-1} d\boldsymbol{\omega} \tag{2}$$

$$= \frac{(-1)^N}{N!} \int_{R^N} ... \int_{R^N} \det{}^2 \|\boldsymbol{\omega}^1, ..., \boldsymbol{\omega}^{N-1}, \boldsymbol{\omega}\| \left(\prod_{n=1}^{N-1} \widetilde{u}(\boldsymbol{\omega}^n) \right) \widetilde{u}\left(\boldsymbol{\omega} - \sum_{n=1}^{N-1} \boldsymbol{\omega}^n \right) e^{i\boldsymbol{\omega}\cdot\mathbf{x}} d\boldsymbol{\omega}^1 ... d\boldsymbol{\omega}^{N-1} d\boldsymbol{\omega}. \tag{3}$$

Here $\|\boldsymbol{\omega}^1, ..., \boldsymbol{\omega}^N\|$ denotes a $N \times N$ matrix comprised of N columnar vectors $\boldsymbol{\omega}^1, ..., \boldsymbol{\omega}^N$. The first factor in the integrand of (3) is a quadratic function of $\boldsymbol{\omega}$, hinting that the l.h.s. of (1) can be transformed into a sum of second derivatives. Indeed, "reverse engineering" of (3) reveals an alternative, *"second-order divergence" form* of (1) in R^N:

$$\frac{1}{N!} \sum_{i_1,...,i_N,j_1,...,j_N} \varepsilon_{i_1...i_N} \varepsilon_{j_1...j_N} \left(u_{x_{i_1}x_{j_1}} ... u_{x_{i_{N-1}}x_{j_{N-1}}} u \right)_{x_{i_N}x_{j_N}} = f. \tag{4}$$

Equivalence of (1) and (4) is now easily established directly: Using the standard rule for differentiation of the products in the l.h.s. of (4), we render it as a sum of products of derivatives of u. Third-order derivatives enter such a product only in pairs of the form $u_{x_{i_p}x_{i_N}x_{j_p}} u_{x_{i_q}x_{j_N}x_{i_q}}$, and hence all products involving third-order derivatives cancel out due to antisymmetry of the tensors $\varepsilon_{i_1...i_N}$. Similarly, all terms involving fourth-order derivatives cancel out. Consequently, each term in the l.h.s. of (4) gives rise to a single term $\varepsilon_{i_1...i_N} \varepsilon_{j_1...j_N} u_{x_{i_1}x_{j_1}} ... u_{x_{i_N}x_{j_N}}$, and hence the sum is the determinant of the Hessian of u, as required.

This form has an interesting consequence. Suppose, for the sake of simplicity, that a space-periodic solution $u \in T^N$ is sought. Let φ be a smooth function with a finite support (in this context such φ are called test functions). Multiplying the MAE by φ and *twice* integrating by parts each term in the sum, we obtain by virtue of (4)

$$\frac{1}{N!} \sum_{i_1,...,i_N,j_1,...,j_N} \varepsilon_{i_1...i_N} \varepsilon_{j_1...j_N} \int_{R^N} u_{x_{i_1}x_{j_1}} ... u_{x_{i_{N-1}}x_{j_{N-1}}} u \, \varphi_{x_{i_N}x_{j_N}} \, d\mathbf{x} = \int_{R^3} f\varphi \, d\mathbf{x}. \tag{5}$$

As usual in the theory of partial differential equations, a weak solution to (1), u, can be defined as a function satisfying the integral identity (5) for any test function φ. (Other definitions of weak solutions to the MAE are natural: generalised solutions obtained by geometric constructions [13] and the so-called viscosity solutions [9]; the two definitions are equivalent [23].) Commonly (e.g., see [17, 25]), only *one* integration by parts is performed in this integral identity. Our form is advantageous in that a lesser regularity of the weak solution is required for (5) to be well-defined. The following argument illustrates this for $N > 2$: For any u from the Sobolev space $W_{N-1}^2(T^N)$ the integrals in the l.h.s. of (5) are well-defined (because by the Sobolev embedding theorem this implies $\nabla u \in L_{2(N-1)}(T^N)$ and hence $u \in L_\infty(T^N)$). By contrast, integrals in the similar identity obtained by just a single integration by parts are *not* well-defined for $u \in W_{N-1}^2(T^N)$. (Note that a viscosity solution for fully nonlinear second-order equations is only required to be continuous [9].)

3

3. The Fourier integral form of the MAE; positivity and bounds of kernels

3.1. Derivation

In the Monge–Ampère–Kantorovich approach to the cosmological reconstruction problems [4, 5], the MAE arises for the potential of the inverse Lagrangian map, in which the function f is a ratio between matter density at the current epoch and at the much earlier epoch of matter-radiation decoupling (at the earlier epoch, the distribution is very close to uniform); thus, $f > 0$.

Here, we consider a problem in which the spatial mean of f is positive. Note that for odd N, if $\langle f \rangle < 0$, the change of the unknown function $u \to -u$ reverses the sign of f and $\langle f \rangle$ admits the desirable sign (here $\langle \cdot \rangle$ denotes space averaging:

$$\langle f \rangle \equiv \lim_{R \to \infty} \frac{1}{|B_R|} \int_{B_R} f(\mathbf{x}) \, d\mathbf{x}, \tag{6}$$

and $|B_R|$ is the volume of the ball $B_R \subset R^N$ of radius R). Suppose that u and f have the same spatial periodicity; integration of (4) over a periodicity cell T^N then yields

$$\int_{T^N} f \, d\mathbf{x} = 0,$$

which is incompatible with $\langle f \rangle \neq 0$. Thus we assume henceforth

$$u = c \left(\frac{|\mathbf{x}|^2}{2} + u' \right), \tag{7}$$

where u' has the periodicity of f and a zero spatial mean (this being just a normalisation), and c is a constant. Substitution of (7) into (1) and integration over a periodicity cell yields

$$c^N = \langle f \rangle. \tag{8}$$

To derive this, note that $u_{x_i x_j} = c(\delta_{ij} + u'_{x_i x_j})$, where δ_{ij} is the Kronecker symbol, and hence the derivatives of u' in the l.h.s. of (1) are present only in $\det \|u'_{x_i x_j}\|$ and a linear combination of minors of smaller sizes of the Hessian of u'. By the algebraic transformation presented in the previous section each such minor can be converted into a sum of second derivatives of products of u' with its second derivatives. Thus the spatial mean of the l.h.s. of (1) over any periodicity cell, and hence the mean defined by (6), does not involve u' and is equal to c^N.

In a more general formulation, we seek a solution (7) satisfying (8) and $\langle u' \rangle = 0$, assuming that f and u' can be represented as Fourier integrals:

$$\nabla^2 u' = \int_{R^N} \widetilde{\eta}(\boldsymbol{\omega}) e^{i \boldsymbol{\omega} \cdot \mathbf{x}} \, d\boldsymbol{\omega}, \quad u' = \int_{R^N} \widetilde{u}'(\boldsymbol{\omega}) e^{i \boldsymbol{\omega} \cdot \mathbf{x}} \, d\boldsymbol{\omega}, \quad \text{where} \quad \widetilde{u}'(\boldsymbol{\omega}) = -\widetilde{\eta}(\boldsymbol{\omega})/|\boldsymbol{\omega}|^2,$$

$$f/c^N = \int_{R^N} \widetilde{f}(\boldsymbol{\omega}) e^{i \boldsymbol{\omega} \cdot \mathbf{x}} \, d\boldsymbol{\omega}.$$

(If the problem, defined by (1), (7) and (8), is considered for space-periodic f and u', integrals in wave vectors in what follows are replaced by the respective Fourier sums.)

Eq. (2) is equivalent to

$$\det \|u'_{x_i x_j}\| = \frac{1}{N!} \int_{R^N} \dots \int_{R^N} \det^2 \left\| \mathbf{i}\boldsymbol{\omega}^1, \dots, \mathbf{i}\boldsymbol{\omega}^{N-1}, \mathbf{i}_{\boldsymbol{\omega} - \sum_{n=1}^{N-1} \boldsymbol{\omega}^n} \right\|$$

$$\times \left(\prod_{n=1}^{N-1} \widetilde{\eta}(\boldsymbol{\omega}^n) \right) \widetilde{\eta} \left(\boldsymbol{\omega} - \sum_{n=1}^{N-1} \boldsymbol{\omega}^n \right) e^{i \boldsymbol{\omega} \cdot \mathbf{x}} \, d\boldsymbol{\omega}^1 \dots d\boldsymbol{\omega}^{N-1} d\boldsymbol{\omega},$$

4

where $\mathbf{i_a}$ denotes a unit vector in the direction of $\mathbf{a}$. Our immediate goal is to derive a similar expression for the terms in (1) of lower orders in u'. The term of order m is

$$\frac{(-1)^m}{N!} \sum_{i_1,...,i_N,j_1,...,j_N} \varepsilon_{i_1...i_N} \varepsilon_{j_1...j_N} \sum_{|\sigma|=m} \left(\prod_{n:\ i_n,j_n \in \sigma} \int_{R^N} \widetilde{u}'(\boldsymbol{\omega}) \omega_{i_n} \omega_{j_n} e^{i\boldsymbol{\omega}\cdot\mathbf{x}}\, d\boldsymbol{\omega} \right) \prod_{n:\ i_n \text{ or } j_n \notin \sigma} \delta_{i_n j_n}$$

(here the sum $\sum_{|\sigma|=m}$ is over all subsets $\sigma \subset \{1,...,N\}$ of cardinality m)

$$= \frac{(-1)^m}{m!} \sum_{1 \le p_1 < ... < p_{N-m} \le N} \int_{R^N} ... \int_{R^N} \left(\sum_{j_1,...,j_m} \varepsilon_{j_1...j_m p_1...p_{N-m}} \prod_{n=1}^{m} \omega_{j_n}^n \right)^2$$

$$\times \left(\prod_{n=1}^{m} \widetilde{u}'(\boldsymbol{\omega}^n) \right) \exp\left(i \sum_{n=1}^{m} \boldsymbol{\omega}^n \cdot \mathbf{x} \right) d\boldsymbol{\omega}^1 ... d\boldsymbol{\omega}^m$$

$$= \int_{R^N} ... \int_{R^N} A_m \left(\mathbf{i}_{\boldsymbol{\omega}^1}, ..., \mathbf{i}_{\boldsymbol{\omega}^{m-1}}, \mathbf{i}_{\boldsymbol{\omega}-\sum_{n=1}^{m-1} \boldsymbol{\omega}^n} \right)$$

$$\times \left(\prod_{n=1}^{m-1} \widetilde{\eta}(\boldsymbol{\omega}^n) \right) \widetilde{\eta}\left(\boldsymbol{\omega} - \sum_{n=1}^{m-1} \boldsymbol{\omega}^n \right) e^{i\boldsymbol{\omega}\cdot\mathbf{x}} d\boldsymbol{\omega}^1 ... d\boldsymbol{\omega}^{m-1} d\boldsymbol{\omega}, \tag{9}$$

where

$$A_m(\mathbf{i}^1, ..., \mathbf{i}^m) \equiv \frac{1}{m!} \sum_{1 \le p_1 < ... < p_{N-m} \le N} M_{p_1...p_{N-m}}^2(\mathbf{i}^1, ..., \mathbf{i}^m) \tag{10}$$

is the sum of squares of all minors of rank m,

$$M_{p_1...p_{N-m}}(\mathbf{i}^1, ..., \mathbf{i}^m) \equiv \sum_{j_1,...,j_m} \varepsilon_{j_1...j_m p_1...p_{N-m}} (\mathbf{i}^1)_{j_1} ... (\mathbf{i}^m)_{j_m},$$

obtained by crossing out rows of numbers $p_1 < ... < p_{N-m}$ from the $N \times m$ matrix

$$\mathcal{M}_m \equiv \left\| \mathbf{i}^1, ..., \mathbf{i}^m \right\|,$$

comprised of m columnar vectors $\mathbf{i}^1, ..., \mathbf{i}^m$.

Therefore, the problem (1) is reduced after the substitution (7) to the system of integral equations, which we call the *Fourier integral form* of the problem (1) and (7):

$$\widetilde{\eta}(\boldsymbol{\omega}) + \sum_{m=2}^{N} \int_{R^N} ... \int_{R^N} A_m \left(\mathbf{i}_{\boldsymbol{\omega}^1}, ..., \mathbf{i}_{\boldsymbol{\omega}^{m-1}}, \mathbf{i}_{\boldsymbol{\omega}-\sum_{n=1}^{m-1} \boldsymbol{\omega}^n} \right)$$

$$\times \left(\prod_{n=1}^{m-1} \widetilde{\eta}(\boldsymbol{\omega}^n) \right) \widetilde{\eta}\left(\boldsymbol{\omega} - \sum_{n=1}^{m-1} \boldsymbol{\omega}^n \right) d\boldsymbol{\omega}^1 ... d\boldsymbol{\omega}^{m-1} = \widetilde{f}(\boldsymbol{\omega}), \tag{11}$$

which is now stated in terms of the Fourier coefficients $\widetilde{\eta}(\boldsymbol{\omega})$ of the Laplacian of the unknown function u'. Equations (11) are valid for all $\boldsymbol{\omega} \ne 0$; the respective equation for $\boldsymbol{\omega} = 0$ is (8).

If the r.h.s. of (1) is zero-mean ($\langle f \rangle = 0$), the MAE admits the Fourier integral form (11) (where $\widetilde{\eta}(\boldsymbol{\omega})$ are now the Fourier coefficients of $\nabla^2 u$), where the l.h.s. is reduced to a single term for $m = N$ in the sum $\sum_{m=2}^{N}$.

3.2. Bounds for the kernels A_m in the Fourier integral form

In this subsection we establish bounds

$$0 \le A_m(\mathbf{i}^1, ..., \mathbf{i}^m) \le \frac{1}{m!}, \tag{12}$$

provided all vectors $\mathbf{i}^s$ have a unit norm. These bounds will play a crucial rôle for our numerical algorithm.

Addition to a column $\mathbf{i}^s$ of any linear combination of columns $\mathbf{i}^{s'}$ for $s' < s$ does not change the value of any minor $M_{p_1...p_{N-m}}(\mathbf{i}^1, ..., \mathbf{i}^m)$. Using the Gram–Schmidt orthogonalisation process, we change all $\mathbf{i}^s$ in $\mathcal{M}_m$ to $\mathbf{j}^s$ such that (i) $\mathbf{j}^1 = \mathbf{i}^1$, (ii) for any s, $\mathbf{j}^s$ differs from $\mathbf{i}^s$ by a linear combination of vectors $\mathbf{i}^{s'}$ for $s' < s$ and thus A_m remains unaltered, (iii) for any s, $\mathbf{j}^s$ is orthogonal to all $\mathbf{j}^{s'}$ for $s' < s$. Hence

$$\mathbf{j}^s = \mathbf{i}_{\mathbf{j}^s} \sin \theta_s,$$

where θ_s is the angle between $\mathbf{i}^s$ and the subspace spanned by $\{\mathbf{i}^{s'} | \ s' < s\}$. Consequently,

$$A_m(\mathbf{i}^1, ..., \mathbf{i}^m) = A_m(\mathbf{j}^1, ..., \mathbf{j}^m) \prod_{s=2}^{m} \sin^2 \theta_s.$$

We denote $\mathcal{M}'_m \equiv \|\mathbf{j}^1, ..., \mathbf{j}^m\|$ and ${}^t\mathcal{M}'_m$ the transpose of $\mathcal{M}'_m$. The identity [26]

$$\sum_{1 \le p_1 < ... < p_{N-m} \le N} M^2_{p_1...p_{N-m}}(\mathbf{j}^1, ..., \mathbf{j}^m) = \det({}^t\mathcal{M}'_m \mathcal{M}'_m) \tag{13}$$

(which does not require orthogonality of $\{\mathbf{j}^1, ..., \mathbf{j}^m\}$) can be easily proved using the formula

$$\det \|a_{ij}\| = \sum_{j_1, ..., j_m} \varepsilon_{j_1...j_m} \prod_{i=1}^{m} a_{ij_i}.$$

We enlarge the set $\{\mathbf{j}^1, ..., \mathbf{j}^m\}$ by vectors $\mathbf{j}^s$ for $s > m$ to a complete orthonormal basis in R^N. Let $\mathcal{U}$ be an orthogonal matrix comprised of the N columnar vectors $\mathbf{j}^1, ..., \mathbf{j}^N$, and $\mathcal{E}$ be the $N \times m$ matrix, whose all entries are 0 except for $\mathcal{E}_{ss} = 1$ for all $1 \le s \le m$. Then $\mathcal{M}'_m = \mathcal{U}\mathcal{E}$ and hence

$$\det({}^t\mathcal{M}'_m \mathcal{M}'_m) = \det({}^t\mathcal{E}{}^t\mathcal{U}\mathcal{U}\mathcal{E}) = \det({}^t\mathcal{E}\mathcal{E}) = \det \mathcal{I}_m = 1,$$

where $\mathcal{I}_m$ is the identity matrix of size m. Consequently,

$$A_m(\mathbf{i}^1, ..., \mathbf{i}^m) = \frac{1}{m!} \prod_{s=2}^{m} \sin^2 \theta_s.$$

This demonstrates that (12) are sharp bounds.

3.3. Solution of the MAE for a weakly fluctuating r.h.s.

If the fluctuating part of the r.h.s. in (1), $f - \langle f \rangle$, is small relative the mean $\langle f \rangle$, (11) suggests that $\widetilde{\eta}(\boldsymbol{\omega}) \approx \widetilde{f}(\boldsymbol{\omega})$ and the nonlinear terms in (11) are small. Then the system can be solved by iteration:

$$\widetilde{\eta}_{K+1}(\boldsymbol{\omega}) = \widetilde{f}(\boldsymbol{\omega}) - \sum_{m=2}^{N} \int_{R^N} ... \int_{R^N} A_m \left(\mathbf{i}_{\boldsymbol{\omega}^1}, ..., \mathbf{i}_{\boldsymbol{\omega}^{m-1}}, \mathbf{i}_{\boldsymbol{\omega} - \sum_{n=1}^{m-1} \boldsymbol{\omega}^n} \right)$$

$$\times \left(\prod_{n=1}^{m-1} \widetilde{\eta}_K(\boldsymbol{\omega}^n) \right) \widetilde{\eta}_K \left(\boldsymbol{\omega} - \sum_{n=1}^{m-1} \boldsymbol{\omega}^n \right) d\boldsymbol{\omega}^1 ... d\boldsymbol{\omega}^{m-1}. \tag{14}$$

Theorem

$1°$. *Suppose positive constants C_0 and C_1 satisfy the inequality*

$$\sum_{m=2}^{N} \frac{(C_0 + C_1)^m}{m!} \le C_1, \tag{15}$$

$$\int_{R^N} |\widetilde{f}(\boldsymbol{\omega})| \, d\boldsymbol{\omega} \le C_0, \tag{16}$$

and $\widetilde{\eta}_0$ satisfies the inequality

$$\int_{R^N} |\widetilde{\eta}_K(\boldsymbol{\omega}) - \widetilde{f}(\boldsymbol{\omega})| \, d\boldsymbol{\omega} \le C_1 \tag{17}$$

for $K = 0$. Then iterates (14) are globally bounded: (17) holds true for all $K > 0$.

$2°$. *Under the same conditions, the following inequalities are satisfied for all $K > 0$:*

$$\int_{R^N} |\widetilde{\eta}_{K+1}(\boldsymbol{\omega}) - \widetilde{\eta}_K(\boldsymbol{\omega})| \, d\boldsymbol{\omega} \le C_2 \int_{R^N} |\widetilde{\eta}_K(\boldsymbol{\omega}) - \widetilde{\eta}_{K-1}(\boldsymbol{\omega})| \, d\boldsymbol{\omega}, \tag{18}$$

$$\max_{\boldsymbol{\omega} \in R^N} |\widetilde{\eta}_{K+1}(\boldsymbol{\omega}) - \widetilde{\eta}_K(\boldsymbol{\omega})| \le C_2 \max_{\boldsymbol{\omega} \in R^N} |\widetilde{\eta}_K(\boldsymbol{\omega}) - \widetilde{\eta}_{K-1}(\boldsymbol{\omega})|, \tag{19}$$

where it is denoted

$$C_2 \equiv \sum_{m=1}^{N-1} \frac{(C_0 + C_1)^m}{m!}.$$

Furthermore, if

$$C_2 < 1, \tag{20}$$

then iterations (14) converge to a solution to (11), which is unique in the ball (17).

The proof is elementary: Inequalities (17)–(19) stem from the bounds (12). By the contraction mapping principle, inequalities (18)–(20) imply convergence of iterations (14) in the norms of $L_1(R^N)$ and $C(R^N)$ in the space of Fourier coefficients, yielding a solution satisfying (17).

Evidently, equations (15) and (20), where equality is assumed instead of the inequalities, have positive solutions for any N. (For instance, $C_0 = \sqrt{3} - 4/3$, $C_1 = 1/3$ for $N = 3$.) If for the chosen values of C_0 and C_1 (15) holds true, $C_2 = 1$ and the inequality (16) is strict, then all conditions of the Theorem become satisfied for a slightly smaller value of C_0.

Note that in view of (20) $C_0 < 1$ and hence (16) implies $f > 0$ everywhere. Consequently, our Theorem is mostly of interest as a statement about convergence of iterations (14). Existence of weak solutions was proved by geometric methods in [13] for the MAE with an arbitrary positive r.h.s. in a compact convex domain. Although our Theorem is significantly weaker because of the strong restrictions on the r.h.s., we have proved it for a non-compact domain — the entire space.

4. The "convolution" form of the MAE

Following the same algebraic ideas, the MAE can be partially "integrated". We again use the Fourier transform of u':

$$u' = \int_{R^N} \widetilde{u}'(\boldsymbol{\omega}) e^{i\boldsymbol{\omega} \cdot \mathbf{x}} \, d\boldsymbol{\omega}, \quad \text{where} \quad \widetilde{u}'(\boldsymbol{\omega}) = (2\pi)^{-N} \int_{R^N} u'(\mathbf{x}) e^{-i\boldsymbol{\omega} \cdot \mathbf{x}} \, d\mathbf{x}.$$

Let

$$\widetilde{\xi}(\boldsymbol{\omega}) \equiv \sqrt{\widetilde{u}(\boldsymbol{\omega})/(2\pi)^N},$$

where

$$\arg(\widetilde{\xi}(\boldsymbol{\omega})) = \arg(\widetilde{u}'(\boldsymbol{\omega}))/2, \quad \text{if} \ \ |\arg(\widetilde{u}'(\boldsymbol{\omega}))| < \pi;$$

$$\arg(\widetilde{\xi}(\boldsymbol{\omega})) = -\arg(\widetilde{\xi}(-\boldsymbol{\omega})), \quad \text{if} \ \ \arg(\widetilde{u}'(\boldsymbol{\omega})) = \pi.$$

That these conditions can be enforced, is elementary for $\boldsymbol{\omega} \neq 0$; the case $\boldsymbol{\omega} = 0$ is not problematic, because $\widetilde{\xi}(0) = 0$. Then

$$\xi(\mathbf{x}) \equiv \int_{R^N} \widetilde{\xi}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega} \cdot \mathbf{x}} \, d\boldsymbol{\omega}$$

is a real-valued function. (For a given u' it is not uniquely defined.)

Let us render the term of order m in u' in the l.h.s. of (1) in the terms of $\xi(\mathbf{x})$ employing the expression (9), (10), (13) and the "identity"

$$(2\pi)^{-N} \int_{R^N} e^{i\boldsymbol{\omega} \cdot \mathbf{x}} \, d\boldsymbol{\omega} = \delta(\mathbf{x})$$

(as usual understood in the sense of generalised functions):

$$\int_{R^N} \cdots \int_{R^N} A_m \left(\mathbf{i}\boldsymbol{\omega}^1, ..., \mathbf{i}\boldsymbol{\omega}^{m-1}, \mathbf{i}_{\boldsymbol{\omega} - \sum_{n=1}^{m-1} \boldsymbol{\omega}^n} \right)$$

$$\times \left(\prod_{n=1}^{m-1} \widetilde{\eta}(\boldsymbol{\omega}^n) \right) \widetilde{\eta}\left(\boldsymbol{\omega} - \sum_{n=1}^{m-1} \boldsymbol{\omega}^n \right) e^{i\boldsymbol{\omega} \cdot \mathbf{x}} \, d\boldsymbol{\omega}^1 ... \, d\boldsymbol{\omega}^{m-1} d\boldsymbol{\omega}$$

$$= (-1)^m \int_{R^N} \cdots \int_{R^N} A_m \left(\boldsymbol{\omega}^1, ..., \boldsymbol{\omega}^m \right) \left(\prod_{n=1}^{m} \widetilde{u}'(\boldsymbol{\omega}^n) \right) \exp\left(i \sum_{n=1}^{m} \boldsymbol{\omega}^n \cdot \mathbf{x} \right) d\boldsymbol{\omega}^1 ... \, d\boldsymbol{\omega}^m$$

$$= \left(-(2\pi)^N \right)^m \int_{R^N} \cdots \int_{R^N} A_m \left(\widetilde{\xi}(\boldsymbol{\omega}^1)\boldsymbol{\omega}^1, ..., \widetilde{\xi}(\boldsymbol{\omega}^m)\boldsymbol{\omega}^m \right) \exp\left(i \sum_{n=1}^{m} \boldsymbol{\omega}^n \cdot \mathbf{x} \right) d\boldsymbol{\omega}^1 ... \, d\boldsymbol{\omega}^m$$

$$= (2\pi)^{-Nm} \int_{R^N} \cdots \int_{R^N} A_m \left(\int_{R^N} \nabla\xi(\mathbf{x}) e^{-i\boldsymbol{\omega}^1 \cdot \mathbf{x}^1} \, d\mathbf{x}^1, ..., \int_{R^N} \nabla\xi(\mathbf{x}) e^{-i\boldsymbol{\omega}^m \cdot \mathbf{x}^m} \, d\mathbf{x}^m \right)$$

$$\times \exp\left(i \sum_{n=1}^{m} \boldsymbol{\omega}^n \cdot \mathbf{x} \right) d\boldsymbol{\omega}^1 ... \, d\boldsymbol{\omega}^m$$

$$= \frac{1}{m!} \int_{R^N} \cdots \int_{R^N} \det\left({}^t \|\nabla\xi(\mathbf{x}^1), ..., \nabla\xi(\mathbf{x}^m)\| \|\nabla\xi(\mathbf{x} - \mathbf{x}^1), ..., \nabla\xi(\mathbf{x} - \mathbf{x}^m)\| \right) d\mathbf{x}^1 ... \, d\mathbf{x}^m.$$

In particular, the linear term ($m = 1$) is $\nabla^2 u' = \int_{R^N} \nabla\xi(\mathbf{x}^1) \cdot \nabla\xi(\mathbf{x} - \mathbf{x}^1) \, d\mathbf{x}^1$.

Therefore, the problem (1) is equivalent, after the substitution (7) and (8), to

$$1 + \sum_{m=1}^{N} \frac{1}{m!} \int_{R^N} \cdots \int_{R^N} \det\left({}^t \|\nabla\xi(\mathbf{x}^1), ..., \nabla\xi(\mathbf{x}^m)\| \|\nabla\xi(\mathbf{x}-\mathbf{x}^1), ..., \nabla\xi(\mathbf{x}-\mathbf{x}^m)\| \right) d\mathbf{x}^1 ... \, d\mathbf{x}^m = \frac{f}{c^N}. \tag{21}$$

We call this integral equation the *"convolution" form* of the MAE. Since it does not involve second-order derivatives, it may prove useful for development of an iterative algorithm for numerical solution of the problem defined by (1), (7) and (8), which involves suitable transformations of the spatial variable.

If the r.h.s. of (1) is zero-mean, the MAE also admits the convolution form (11), where $c = 1$, the l.h.s. is reduced to a single term for $m = N$ in the sum $\sum_{m=1}^{N}$, and $\widetilde{u}'(\boldsymbol{\omega})$ in the definition of $\widetilde{\xi}(\boldsymbol{\omega})$ denotes the Fourier transform of u.

5. A test problem with a cosmological flavour

One important area of application of the MAE is the reconstruction of the dynamical history of the Universe from present observations of the spatial distribution of masses (galaxies, clusters, including their dark-matter components). Let us briefly recall the background. Peebles was the first to propose that, from the sole knowledge of the current positions of galaxies (from the Local Group which includes our own galaxy) without knowledge of their (proper) velocities, reconstruction of the full dynamical history is a meaningful goal [27]. Indeed, the very strong constraint on the distribution of masses at the epoch of decoupling — which has to be almost uniform — makes reconstruction a possibly well-posed two-point boundary problem that can then be solved by variational techniques. In fact, using convexity techniques, a theorem of uniqueness of reconstruction was proved in [5], using the Euler–Poisson equations, which describes the dynamics of matter on sufficiently large scales (of the order of a few million light years). In practice, reconstruction on such scales was done so far via the Monge–Ampère–Kantorovich (MAK) method [4] (see also [5, 6]) which assumes that the Lagrangian map from initial to current mass locations is the gradient of a convex potential. This holds exactly for the Zel'dovich approximation [28] (in which the equation for the velocity of matter reduces to the inviscid Burgers equation with a zero r.h.s., and hence the peculiar velocities remain constant along trajectories of point masses) and also for its refinement, the first-order Lagrangian perturbation approximation [29]. It is then easy, using mass conservation, to derive a MAE for the potential of the inverse Lagrangian map (the latter is the Legendre–Fenchel transform of the potential of the direct map). By a theorem of Brenier [30] the MAE becomes a Monge–Kantorovich mass transportation problem with quadratic cost which, after discretisation, can be solved by optimisation techniques (see [5] for details). The r.h.s. of the MAE is equal to the ratio of densities at the present and initial positions. Thus, if the initial density was not uniformly distributed, we would have to solve the MAE with the r.h.s. depending on the gradient of the unknown potential. However, at the epoch of decoupling (about 380 thousand years after the Big Bang) the density of matter was close to uniform, which implies a significant simplification of the MAE to be solved: its r.h.s. is known.

Here we explore for the first time the possibility of directly solving the three-dimensional MAE without discretisation on a toy model with a cosmological flavour.

We assume that the dimension of space is $N = 3$, and the r.h.s. of the MAE has the following structure:

$$f = \delta^{-3} \sum_{g=1}^{G} f^{(g)} \left(\frac{\mathbf{r} - \mathbf{r}^{(g)}}{\delta} \right). \tag{22}$$

The function (22) describes mass distribution for G "objects", which are "localised", if the value of the parameter δ is small compared to the distance between objects. In the cosmological context galaxies or clusters of galaxies can be regarded as such objects; then $f^{(g)}$ describes the *total* distribution of mass of both visible matter and dark matter, in which the galaxies are embedded. Although observations attest, that the distribution of visible matter in galaxies is clearly discontinuous, it is astrophysically sound to expect that the distribution of all types of matter is smooth due to the prevailing smoothness of the dark matter. We assume that objects have density distributions with a Gaussian shape:

$$f^{(g)}(\mathbf{r}) = \frac{m^{(g)}}{(\sigma^{(g)} \sqrt{\pi})^3} \exp(-|\mathbf{r}/\sigma^{(g)}|^2). \tag{23}$$

The g-th object of mass $m^{(g)} > 0$ is located at $\mathbf{r}^{(g)}$. All $m^{(g)}$, $\sigma^{(g)}$ and $\mathbf{r}^{(g)}$ are independent of the small parameter $\delta > 0$.

We will be seeking a solution (7) with a space-periodic u' to the Monge–Ampère problem (1), where a space-periodic r.h.s. $\widehat{f}$ is the sum of "clones" of (22) over all periodicity cells. Without loss of generality we can assume that the total mass is normalised:

$$\int f(\mathbf{r}) \, d\mathbf{r} = \sum_{g=1}^{G} m^{(g)} = 1;$$

thus (8) implies $c = 1$ in (7).

It is instructive to consider particular solutions to this Monge–Ampère problem.

5.1. An exact solution to the MAE for a spherically symmetric mass distribution

In spherical coordinates centred at $\mathbf{r}^{(g)}$, (1) becomes, for spherically symmetric u and f,

$$\rho^{-2} \frac{\partial^2 u}{\partial \rho^2} \left(\frac{\partial u}{\partial \rho} \right)^2 = \delta^{-3} f,$$

where $\rho = |\mathbf{r} - \mathbf{r}^{(g)}|$. This equation has an obvious solution

$$u(\rho, \delta) = \int_0^\rho \left(3 \int_0^{r'/\delta} r^2 f(r) \, dr \right)^{1/3} dr', \tag{24}$$

whose first derivatives are uniformly bounded:

$$\frac{\partial u}{\partial x_i} = \frac{x_i}{\rho} \left(3 \int_0^{\rho/\delta} r^2 f(r) dr \right)^{1/3} = O(\delta^0);$$

the second derivatives are $O(\delta^{-1})$:

$$\frac{\partial^2 u}{\partial x_i \partial x_m} = \delta^{-1} \frac{x_i x_m}{\rho^2} \left(\frac{\rho}{\delta} \right)^2 f\left(\frac{\rho}{\delta} \right) \left(3 \int_0^{\rho/\delta} r^2 f(r) dr \right)^{-2/3} + \left(\delta_m^i - \frac{x_i x_m}{\rho^2} \right) \frac{1}{\rho} \left(3 \int_0^{\rho/\delta} r^2 f(r) dr \right)^{1/3}.$$

This estimate follows from the following inequalities:
- $|x_i x_m / \rho^2| \le 1$;
- for $\rho < \delta$, $f(\rho/\delta)$ is uniformly bounded and

$$\underline{c}(\rho/\delta)^3 \le \int_0^{\rho/\delta} r^2 f(r) dr \le \overline{c}(\rho/\delta)^3$$

for some positive constant $\underline{c}$ and $\overline{c}$;
- for $\rho \ge \delta$, $(\rho/\delta)^2 f(\rho/\delta)$ is uniformly bounded and

$$0 < \underline{c}' \le \int_0^{\rho/\delta} r^2 f(r) dr \le \overline{c}'$$

for some constant $\underline{c}'$ and $\overline{c}'$.

Moreover, if ρ is larger than a positive constant, $f(\rho/\delta) = o(\delta^3)$, and hence all second derivatives of u are uniformly bounded outside any sphere of a fixed radius.

The properties of the solution discussed in this subsection are implied just by a fast decay of f at infinity and the spherical symmetry of the r.h.s.; the assumption that the profile is Gaussian is unnecessary.

5.2. A one-cell solution for $G > 1$ spherically symmetric localised objects

A solution for $G > 1$ objects can be expected to admit a power series expansion

$$u(\mathbf{r}) = \sum_{n=0}^\infty u_n(\mathbf{r}, \delta) \delta^n,$$

where (by analogy with (24)) each $u_n(\mathbf{r}, \delta)$ and its first derivatives are $O(\delta^0)$, and the second derivatives are $O(\delta^{-1})$.

Naively one could expect interaction of localised objects to be asymptotically unimportant, and hence to obtain in the leading order the sum of individual one-object solutions:

$$u_0(\mathbf{r}) = \sum_{g=0} u^{(g)}(|\mathbf{r} - \mathbf{r}^{(g)}|, \delta). \qquad (25)$$

Here $u^{(g)}(|\mathbf{r} - \mathbf{r}^{(g)}|, \delta)$ is the one-object solution (24) for the g-th object. Let us inspect, to what extent such a conjecture might be true.

Consider a neighbourhood of an object γ. In the leading order, the l.h.s. of (1) is $\det \|u_{0\ x_i x_j}\|$. Substitution of (25) into the determinant yields a sum of triple products of second derivatives of various $u^{(g)}$. As shown in the previous subsection, for $g \neq \gamma$ each second derivative of $u^{(g)}$ is $O(\delta^0)$ in this neighbourhood (the distance between the objects g and γ is a fixed positive constant). Thus, only triple products of second derivatives of $u^{(\gamma)}$ contribute terms of the leading order, $O(\delta^{-3})$, the one of the l.h.s. These products constitute $\det \|u_{x_i x_j}^{(\gamma)}\|$ and thus match the term $\delta^{-3} f^{(\gamma)}$ in the r.h.s. Nonlinear interaction of pairs of solutions for individual objects is at a lower order, $O(\delta^{-2})$, and that of triplets of one-object solutions - still weaker, $O(\delta^{-1})$. Hence, after the finite sum (24) is substituted into the MAE, the highest order terms do cancel out, and in this respect the naive conjecture is confirmed.

Nevertheless, the one-cell solution (25) does not represent the leading order term of the solution to the problem at hand. To see this, note that at large distances (25) exhibits a linear growth in $|\mathbf{r}|$, and not a quadratic one, as the form of the solution (7) requires. Furthermore, after subtraction of the quadratic profile $c|\mathbf{r}|^2$, the remaining part u' must be space-periodic, which is clearly not the case for (25). We cannot use (25) to construct a global solution conformant with this periodicity requirement by the procedure, used to obtain the periodic r.h.s. $\widehat{f}$ from the individual density profiles (22), because the sum of periodically distributed "clones" of one-cell solutions (25) is infinite. Moreover, for such hypothetical space-periodic sum there would be infinitely many pairwise interactions between objects yielding products of the order $O(\delta^{-2})$; hence, even if the sum of such products is finite, it will not necessarily remain $O(\delta^{-2})$.

We observe that the pairwise interaction does not become weaker when the distance between the interacting objects grows. Thus, even in the high-contrast limit the solution that we are seeking does not reduce to individual interactions between objects; their interaction is essentially collective, which makes it hard to predict the structure of the solution.

6. Computation of a solution to the MAE with a space-periodic r.h.s.

In this section we present two iterative algorithms for numerical solution of the space-periodic Monge–Ampère problem. One version (AICDM), employing numerical improvement of convexity and discrepancy minimisation stabilising the iterative process, is suitable for computation of solutions for everywhere positive right-hand sides f (see Subsection 6.4). Another one (ACPDM), involving continuation in parameter and discrepancy minimisation, does not require this condition to be satisfied (see Subsection 6.2). They both rely on the basic algorithm for iterative solution of a fixed point problem for the MAE (see Subsection 6.1). Test applications of the two algorithms are considered in Subsections 6.3 and 6.4. We assume $\langle f \rangle \neq 0$, however, reformulation of ACPDM for the case of a zero-mean space-periodic r.h.s. is straightforward.

6.1. A basic solver (FPAR)

If the amplitude of fluctuation of f is small compared to the mean, — more precisely, if (16) is satisfied, — the Theorem (see Subsection 3.3) establishes existence of a solution to (1), that is a perturbation of $c|\mathbf{x}|^2/2$. If, in addition, (20) holds true, the Theorem guarantees that iterations defined by (14) converge to the Laplacian of the perturbation; this offers an algorithm for numerical solution of the MAE. If either of the conditions (16) or (20) are violated (which is the practically interesting case), iterations (14) do not necessarily converge. Different algorithms are necessary, taking this into account.

Suppose the kernels A_m in (11) are frozen and take constant values a_m, respectively. Then the new system of equations (11) can be easily solved, since it is the Fourier form of the polynomial equation

$$\sum_{m=0}^{N} a_m \eta^m = f/\langle f \rangle$$

in the physical space. This observation suggests the following algorithm. We express (11) in the terms of $\eta = \nabla^2 u'$ as

$$\sum_{m=0}^{N} a_m \eta^m = f/\langle f \rangle + F(\eta), \qquad (26)$$

where

$$F(\eta) \equiv 1 + \sum_{m=1}^{N} a_m \eta^m - \det \| \nabla^{-2}(\eta_{x_i x_j}) + \delta_{ij} \|,$$

and implement iterations

$$\sum_{n=0}^{N} a_m \eta_K^m = f/\langle f \rangle + F(\eta_{K-1}) \qquad (27)$$

in the physical space.

The applicability of this algorithm depends crucially on the choice of the coefficients a_m. In view of the positivity and the bounds for the kernels (12), we impose

$$0 \leq a_m \leq \frac{1}{m!} \quad \text{for} \;\; m > 0.$$

Furthermore, it seems practical to set

$$a_1 = 1, \qquad a_m = \frac{1}{2m!} \quad \text{for} \;\; m > 1, \qquad (28)$$

because for $a_1 = 1$ the linear in u' term is treated exactly,

$$F(\eta) = \sum_{m=2}^{N} \int_{R^N} ... \int_{R^N} \left(A_m(\mathbf{i}\boldsymbol{\omega}^1, ..., \mathbf{i}\boldsymbol{\omega}^m) - a_m \right) \left(\prod_{n=1}^{m} \widetilde{\eta}(\boldsymbol{\omega}^n) \right) \exp\left(i \sum_{n=1}^{m} \boldsymbol{\omega}^n \cdot \mathbf{x} \right) d\boldsymbol{\omega}^1 ... d\boldsymbol{\omega}^m \qquad (29)$$

and the median values (28) of A_m for $m > 1$ minimise the ranges of the kernels

$$A_m(\mathbf{i}\boldsymbol{\omega}^1, ..., \mathbf{i}\boldsymbol{\omega}^m) - a_m$$

in (29). Since for the coefficients (28)

$$|A_m(\mathbf{i}\boldsymbol{\omega}^1, ..., \mathbf{i}\boldsymbol{\omega}^m) - a_m| \leq a_m, \quad m \geq 2$$

the l.h.s. of (27) "captures" the nonlinear behaviour of the l.h.s. of (11), the algorithm has chances to converge. However, for the extreme values of A_m the values of the kernels in (29) are as large as the respective medians, and therefore convergence of iterations (27), (28) is not guaranteed.

At least two strategies can be proposed for setting the value of a_0 (which is a free parameter in the sense that (11) is not required to be satisfied for $\boldsymbol{\omega} = 0$):

1°. At each iteration a_0 is tuned, so that $\langle \eta \rangle = 0$ (which holds true for $\eta = \nabla^2 u'$).

2°. We set $a_0 = 0$.

Note that although in the subspace of space-periodic functions the inverse Laplacian is defined for zero-mean scalar fields only, $\langle \eta_K \rangle$ is not required to vanish, because the mean is removed by differentiation in the Hessian before the inverse Laplacian is evaluated.

12

For any odd N, the equation (27) in η_K with the coefficients (28) has a unique root for any r.h.s. To check this, it is enough to establish that the derivative $D_N(\eta)$ of the l.h.s. of (27) is positive for any η. For the choice of coefficients (28),

$$D_N(\eta) - D'_N(\eta) = \frac{1}{2}\left(1 + \frac{\eta^{N-1}}{(N-1)!}\right).\tag{30}$$

At a minimum $D'_N(\eta) = 0$ and hence (30) implies that at the minimum $D_N > 0$, proving monotonicity of the l.h.s. of (27). It can be shown similarly, that for any even N the number of roots is 0 or 2. Thus, the algorithm is guaranteed to be applicable for odd N only.

It can be proved, that the Theorem also applies for iterations (27) (with the same constant C_1 bounding the same norm of solutions).

In an application to a test problem inspired by cosmology (see Subsection 6.3), this algorithm, which we call FPAR (*Fixed Point Algorithm for the Regular part of the MAE*) produces a sequence of iterations, initially converging, but subsequently blowing up: linear instability sets in, and hence the respective unstable mode must be removed.

6.2. A more advanced solver (ACPDM)

The behaviour of the basic algorithm FPAR suggests that it should be embedded as an engine within a more advanced algorithm. In this subsection we present such an advanced solver, ACPDM (*Algorithm with Continuation in a Parameter and Discrepancy Minimisation*).

Consider a generalisation of (26):

$$Q(\eta; p) = 0,\tag{31}$$

where

$$Q(\eta; p) \equiv \sum_{m=0}^{N} a_m \eta^m - f/\langle f\rangle - pF(\eta).$$

Here the coefficients (28) are assumed, and the new parameter p is confined to the interval $[0, 1]$. For $p = 0$, (31) is just a set of polynomial equations of degree N; for $p = 1$ it reduces to (26) which is equivalent to (1). Continuation in the parameter p is implemented: (31) is solved for a set of values p_j, increasing from 0 to 1, and an initial approximation of the solution for $p = p_j$ is obtained by polynomial extrapolation of solutions for all $p_{j'} < p_j$. (Numerical extrapolation requires performing quadruple precision computations, if the number of nodes exceeds roughly a dozen.) For any p, solution of (31) involves iterations similar to (27):

$$\sum_{m=0}^{N} a_m \eta_K^m = f/\langle f\rangle + pF(\eta_{K-1}).\tag{32}$$

The r.m.s. discrepancy

$$d(\eta) \equiv \sqrt{\left\langle \left(Q(\eta; p) - \langle Q(\eta; p)\rangle\right)^2\right\rangle}$$

is used in the termination condition. Note that $\langle Q(\eta; 1)\rangle = 0$ and hence

$$d(\eta) \equiv \sqrt{\left\langle \left(\det\|\nabla^{-2}\eta_{x_i x_j} + \delta_{ij}\| - f/\langle f\rangle\right)^2\right\rangle}$$

for $p = 1$; in particular, $d(\eta) = 0$ for $p = 1$, if and only if for $u' = \nabla^{-2}\eta$ and the normalisation (8) the field (7) is a solution to the MAE.

Let $(\cdot, \cdot)$ denote a scalar product and $\|\cdot\|$ the induced norm of a scalar field: $\|\mathbf{v}\| \equiv \sqrt{(\mathbf{v}, \mathbf{v})}$. In the test runs reported in the next subsection, the scalar product of the functional Lebesgue space $L^2(T^3)$ has been assumed:

$$(\mathbf{u}, \mathbf{v}) = \int_{T^3} u(\mathbf{x})v(\mathbf{x})d\mathbf{x}.\tag{33}$$

Other products (with different weight functions introduced in the above integral) have been also considered (see Subsection 6.4).

If η' and η'' are approximate solutions to the generalised MAE (31), then

$$Q(\eta'; p) - Q(\eta''; p) = A(\eta' - \eta'') + O(\|\eta' - \eta''\|^2), \tag{34}$$

where A is the linearisation of (31) around the solution. Consequently, the concept of minimisation of the residual in Krylov spaces and approaches for its realisation can be borrowed from solvers for linear problems (such as the Generalised Conjugate Gradients Method [31]). ACPDM involves sequences of stabilised iterations described below, which exploit this concept. To make such an iteration, the following data is required: an approximate solution η_K, and two sets of S scalar fields, $v_s(\mathbf{x})$ and $w_s(\mathbf{x})$, $0 \le s \le S$, where $0 \le S \le S_{\max}$ (as well as some other quantities computed at the previous stabilised iteration), and $S_{\max}$ is a parameter of the algorithm. It is supposed that all v_s are mutually orthogonal with respect to the scalar product $(\cdot, \cdot)$, and

$$v_s = A w_s + O(\|w_s\|^2) \tag{35}$$

for small w_s.

A *sequence of stabilised iterations of ACPDM* is initialised using the current approximation η_0 by computing $\eta_1 = \eta_1'$ as a solution to (32) for $K = 1$, and setting $S = 0$ (i.e. the sets $w_s(\mathbf{x})$ and $v_s(\mathbf{x})$ are empty). For $K > 1$, a *stabilised iteration of ACPDM* consists of the following steps:

i. At each grid point in the physical space solve equation (32):

$$\sum_{m=0}^{N} a_m (\eta_K')^m = f/c^N + p F(\eta_K).$$

ii. Compute $Q(\eta_K')$.

iii. Orthogonalise $v' \equiv Q(\eta_K') - Q(\eta_{K-1}')$ to all v_s, $1 \le s \le S$, with respect to the scalar product $(\cdot, \cdot)$, and set

$$v_{S+1} = v' - \sum_{s=1}^{S} \frac{(v', v_s)}{(v_s, v_s)} v_s, \qquad w_{S+1} = \eta_K' - \eta_{K-1}' - \sum_{s=1}^{S} \frac{(v', v_s)}{(v_s, v_s)} w_s.$$

iv. Compute

$$\eta_{K+1} = \eta_K' - \sum_{s=1}^{S+1} \frac{(Q(\eta_K'), v_s)}{(v_s, v_s)} w_s.$$

v. If $S < S_{\max}$, increase S by 1; otherwise (i.e., if $S = S_{\max}$), discard v_1 and w_1, and decrease by 1 the indices s of the remaining $S_{\max}$ fields v_s and w_s.

vi. Compute the r.h.s. of (32) for η_{K+1} substituted in place of η_{K-1}, the field $Q(\eta_{K+1})$ and $d(\eta_{K+1})$. If $d(\eta_{K+1})$ is less than a given small threshold, then η_{K+1} is the desired approximate solution and computation for the present p is finished. If $\|Q(\eta_{K+1})\| > \|Q(\eta_K)\|$, then the current sequence is terminated.

A few comments are in order. Clearly, v_{S+1} and w_{S+1} obtained in step *iii*, possess the required properties: v_{S+1} is $(\cdot, \cdot)-$orthogonal to all v_s for $s \le S$, and (35) holds for v_{S+1} and w_{S+1} by virtue of (34). The coefficients in the sum computed in step *iv* minimise the discrepancy $\|Q(\eta_K' - \sum_{s=1}^{S} q_s w_s)\|$, assuming all $q_s w_s$ are small and hence their quadratic contributions are negligible. The minimisation plays a dual rôle: on the one hand, it stabilises basic iterations (32), removing the instability modes as soon as they become substantial; on the other, it significantly increases the efficiency of the algorithm (up to a factor 20 compared to other algorithms relying on iterations (32)). The assumption that nonlinear terms are small can be incorrect; also, as a result of accumulation of the neglected nonlinear terms after a number of steps, at some stage v_s can cease to approximate $A w_s$ accurately enough. If the inequality $\|Q(\eta_{K+1})\| > \|Q(\eta_K)\|$ is found to hold true in step *vi*, we interpret this as an indication that the adverse effect of nonlinearity has become significant, and then the algorithm breaks the current sequence. In order to reduce the adverse influence of nonlinearity, a small number $S_{\max}$ may be chosen; in our runs, $S_{\max} = 5$.

Now we can assemble the algorithm from the building blocks, discussed above. ACPDM performs continuation in the parameter p. For a given p, it starts by carrying out basic iterations (32). As soon as convergence slows down (we have used the condition $d^2(\eta_K) > d^2(\eta_{K-1})/2$), the algorithm switches

to perform a sequence of stabilised iterations. If the sequence terminates in step vi because nonlinear effects became significant, the algorithm proceeds by performing basic iterations (32). Usually it takes a small number of them for the instability to set in, and as soon as ACPDM detects that the inequality $d^2(\eta_K) > \alpha d^2(\eta_{K-1})$ holds true, it starts a new sequence. Here α is a parameter, which can slightly exceed 1 (in order to allow the transients to die off and the dominant instability modes to set in, so that the latter could be efficiently removed in subsequent stabilised iterations); in our test runs we have used $\alpha = 1.05$.

6.3. Application to a test problem inspired by cosmology

We have chosen to employ a problem of the kind discussed in Section 5 as a test-bed for our algorithm, because cosmological applications of the MAE are probably the most important ones. The positiveness of the r.h.s. (implying convexity of solutions [13]) is not required for application of the algorithm.

We are seeking a solution (7) with a space-periodic u' to the Monge–Ampère problem (1), where a space-periodic r.h.s. $\widehat{f}$ is a sum of "clones" of (22) over all periodicity cells. Our poor man's Universe involves $G = 3$ objects in the periodicity cell $T^3([0,1]^3)$, centred at the origin: $-1/2 \le x_i \le 1/2$. They are described by the following parameter values in (23):

$$m^{(1)} : m^{(2)} : m^{(3)} = \frac{1}{18} : \frac{1}{27} : \frac{1}{32},$$

$$\delta = 1, \qquad \sigma^{(1)} = \sigma^{(2)} = \frac{1}{6}, \quad \sigma^{(3)} = \frac{1}{8},$$

$$\mathbf{r}^{(1)} = \frac{1}{4}(-1,1,-1), \quad \mathbf{r}^{(2)} = \frac{1}{4}(1,-1,-1), \quad \mathbf{r}^{(3)} = \frac{1}{4}(-1,-1,1). \tag{36}$$

Noting that $\widehat{f}(\mathbf{x})$ achieves its maximum at $\mathbf{r}^{(3)}$ and its minimum at $\frac{1}{4}(1,1,1)$, we can estimate the contrast number of the problem as

$$\frac{16}{4\left(12\exp(-18) + 8\exp(-18) + 16\exp(-32)\right)} \approx 1.313 \times 10^7$$

(the factor 4 in the denominator accounts for the replicas of (23), located in the neighbour periodicity boxes at a distance $1/\sqrt{2}$).

Iterations defined by (32) have been performed on a uniform grid comprised of 64^3 points in the periodicity cell $T^3([0,1]^3)$. We have evaluated $\det \|u'_{x_i x_j} + \delta_{ij}\|$ by the pseudospectral method with dealiasing (for $N = 3$ this requires computation of the derivatives $u'_{x_i x_j}$ in the physical space on the twice finer grid involving 128^3 points). Apparently dealiasing does not play any important rôle in these computations; this is consistent with the fast decay of the energy spectrum of $\eta = \nabla^2 u'$ by 11 orders of magnitude.

We have tested both strategies for choosing a_0 (see Subsection 6.1), as well as a hybrid strategy, in which $a_0 = 0$, but after solving (27) we reset $\eta_K := \eta_K - \langle \eta_K \rangle$. It turns out that the hybrid strategy and 1° are very close in the number of iterations necessary to obtain a solution to the MAE to the same accuracy $d(\eta_K) < 10^{-10}$. However, this implies that the hybrid strategy is several times more efficient in the terms of CPU time, since it requires just one evaluation of solutions to (27), while evaluation of a_0 following strategy 1° requires several such evaluations. The strategy 2° has proved to be the most efficient one, both in the terms of CPU time and the number of required iterations. (The same nodes p_j were used in all the test runs.) In the remaining part of the subsection we discuss convergence of the advanced method with $a_0 = 0$.

ACPDM performs a combination of basic iterations (32) and stabilised iterations, which have different computational costs. The most time-consuming operation, determining the cost, is computation of the determinant of the Hessian in $F(\eta)$. There are two such operations in a stabilised iteration, and a single one in a basic iteration. Consequently, the former is approximately twice longer than the latter. To enable an accurate comparison of performance of various versions of our code, we report the number of evaluations of determinant in each run (a run is terminated when the obtained approximate solution η satisfies the accuracy requirement $d(\eta) = 10^{-10}$). Note however, that the main bulk of computations is performed by
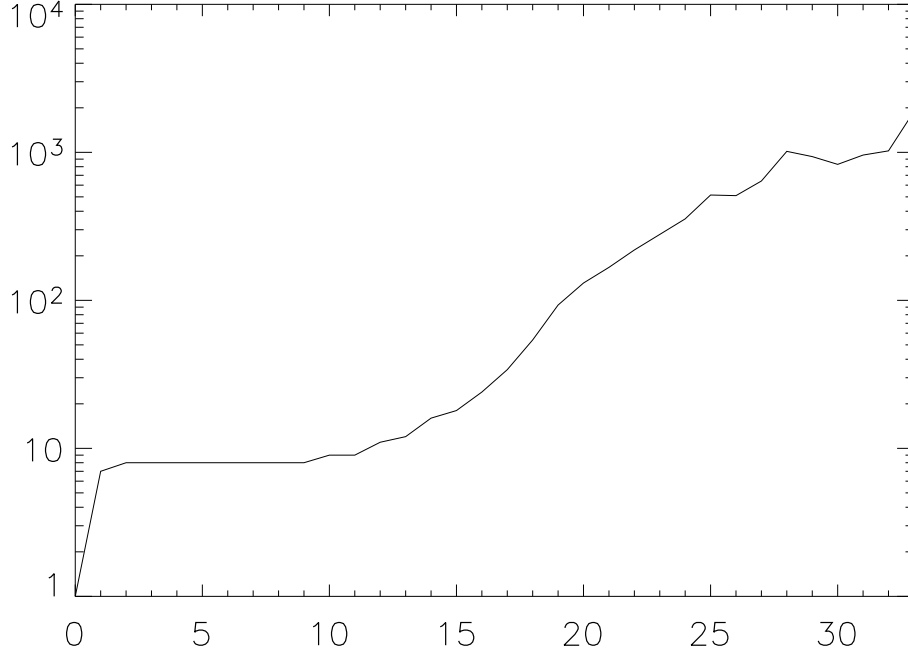
Figure 1: Number of evaluations of the determinant of the Hessian (vertical axis, logarithmic scale) performed by ACPDM in successive computations of approximate solutions $\eta(p_j)$ to the generalised MAE (31), satisfying $d(\eta(p_j)) < 10^{-10}$, for numerical solution of the test MAE (31) (see Subsection 6.3) Horizontal axis: the index j numbering consecutive nodes p_j in the mesh (37). The initial approximation for a p_j is obtained by the polynomial extrapolation of solutions for $p_{j'}$ with $j' < j$.

doing stabilised iterations; the number of basic iterations is usually below 1% of the number of stabilised ones.

In all runs, the initial iteration is a solution to (31) for $p = 0$. If ACPDM is applied to this field for $p = 1$, iterations quickly become chaotic and cease to converge (this has necessitated to include into the algorithm continuation in the parameter p). Initially, ACPDM has been applied for uniformly distributed nodes $p_j = j/J$, $j = 0, ..., J-1$ for $J = 20$, for which the algorithm has shown a remarkably fast convergence (see Fig. 1). A polynomial 20-node extrapolation yields an approximate solution to our test problem to the accuracy $d(\eta) = 0.47 \times 10^{-2}$ and $d_\infty(\eta) = 0.02$, where

$$d_\infty(\eta) \equiv \max_{T^3} \left| \det \|u'_{x_i x_j} + \delta_{ij}\| - \widehat{f}/\langle\widehat{f}\rangle \right|.$$

When it is used as an initial approximation for a run for $p = 1$, convergence is slow and the pattern of convergence is erratic. It takes $38\,685$ evaluations of the determinant of the Hessian for the two discrepancy norms to decrease to 0.99×10^{-9} and 0.39×10^{-7}, respectively. At this stage convergence stalls. We have therefore got into a local minimum of $d(\eta)$, out of which no exit can be found; we will refer to it as a spurious minimum solution.

A better approximation to the solution (7) to our test problem is obtained by adding more nodes near the right endpoint $p = 1$. We have chosen to add $J' = 13$ nodes, the complete p-mesh being

$$p_j = j/J, \quad j = 0, ..., J-1; \qquad p_{J+j-1} = 1 - (2^j J)^{-1}, \quad j = 1, ..., J'; \qquad p_{J+J'} = 1 \qquad (37)$$

(information on convergence at the new nodes p_{J+j} is also included in Fig. 1). Initial approximations at each new p_{J+j} become more and more accurate in the interval $0 < p \le 0.55$ ($j = 11$), then the discrepancy $d(\eta)$ of the initial approximations starts growing and admits the maximum 0.52×10^{-3} for $p = 0.9875$ ($j = 21$), and subsequently decreases again. The rate of convergence progressively falls down as p_{J+j} approaches 1: the number of evaluations of the determinant of the Hessian yielding solutions of the

16

Table 1: Number of evaluations (NE) of the determinant of the Hessian, performed by AICDM in the course of computation of approximate solutions (7) of varying accuracy to the test MAE with the r.h.s. defined by (22), (23) and (36), discrepancy $d_\infty(\eta)$ for these solutions, and wallclock duration of runs (DR, seconds) for computation of approximate solutions to the test MAE by EAICDM.

$d(\eta)$	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
$d_\infty(\eta)$	0.64×10^{-2}	0.86×10^{-3}	1.02×10^{-3}	0.23×10^{-4}	0.28×10^{-5}	0.39×10^{-6}	0.36×10^{-7}	0.41×10^{-8}
NE	64	126	251	439	1 018	1 515	2 030	2 653
DR	3.38	4.69	11.17	42.7	81	203	435	774

Table 2: Number of evaluations (NE) of the determinant of the Hessian, performed by AICDM with the use of various weight functions (38) in the scalar product $(\cdot,\cdot)$. All runs are terminated as soon as the accuracy $d(\eta) < 10^{-10}$ is obtained; $d(\eta) = 0.66 \times 10^{-10}$ for $q = -1$, $d(\eta) = 0.95 \times 10^{-10}$ for $q = -3/4$ and $d(\eta) = 1.0 \times 10^{-10}$ in all remaining cases.

q	-1	$-3/4$	$-1/2$	$-1/4$	0	$1/4$
NE	3 169	2 619	2 465	3 743	2 653	2 571
$d_\infty(\eta)$	0.09×10^{-8}	0.14×10^{-8}	0.21×10^{-8}	0.62×10^{-8}	0.41×10^{-8}	0.39×10^{-8}

q	$1/2$	$3/4$	1	$5/4$	$3/2$	$7/4$	2
NE	2 568	2 643	2 523	2 489	2 481	2 505	2 526
$d_\infty(\eta)$	0.37×10^{-8}	0.39×10^{-8}	0.36×10^{-8}	0.38×10^{-8}	0.39×10^{-8}	0.41×10^{-8}	0.37×10^{-8}

desired accuracy $d(\eta(p_{J+j})) < 10^{-10}$, markedly increases. However, ACPDM does not stall any more. A polynomial extrapolation involving the 33 nodes delivers an approximate solution u' for $p = 1$ to the accuracy $d(\eta) = 0.78 \times 10^{-7}$ and $d_\infty(\eta) = 0.20 \times 10^{-5}$, and after further 1 885 evaluations of the determinant ACPDM yields an approximate solution with the two discrepancy norms down to 10^{-10} and 0.26×10^{-8}, respectively. In total, 9 814 evaluations of the determinant of the Hessian are involved in computations on the mesh (37) providing the solution to the MAE.

The geometry of the obtained solution — isosurfaces of u' and $\nabla^2 u'$ — is shown in Figs. 2 and 3, respectively. The structures disclosed by the isosurfaces of $\nabla^2 u'$ (Fig. 3) are clearly associated with the three objects incorporated into the r.h.s. (22) of the test problem. The figures also reveal a subtle interaction of the objects along the lines connecting their centres (see in Fig. 3 the tube-like structures, which connect the regions of higher values of $\nabla^2 u'$ encompassing the centres of objects).

Figures 4 and 5 display the structure of the field $\nabla u'$, determining the displacement of mass in the test Universe. It turns out that u' possesses hidden symmetries, namely, mirror reflection symmetries about any plane that is parallel to a coordinate plane and contains a pair of objects. (Deliberately, our computations are not sped up by exploiting these symmetries.) Because of the symmetries, the component of $\nabla u'$ normal to such a plane vanishes on this plane. To illustrate the behaviour of the components of the gradient not shown in Fig. 4, we present in Fig. 5 their isolines on coordinate planes, which pass through the centre of the periodicity cube $T^3([0,1]^3)$ and hence are displaced from the planes shown in Fig. 4 by a quarter of period.

6.4. A solver for MAE with an everywhere positive r.h.s. (AICDM)

Clearly, it is desirable to avoid the refinement of solutions for nodes $0.95 < p_j < 1$, which has proved necessary in application of ACPDM to our test problem. Comparison of the spurious minimum solution, obtained with the 20 nodes p_j, and the more accurate one is instructive. The maximum discrepancy between the two approximations of $\eta = \nabla^2 u'$ for these solutions, equal to 0.132, is attained at the minimum of the r.h.s., $(1/4, 1/4, 1/4)$, and all points, where the discrepancy is larger than 0.02 are located within the distance
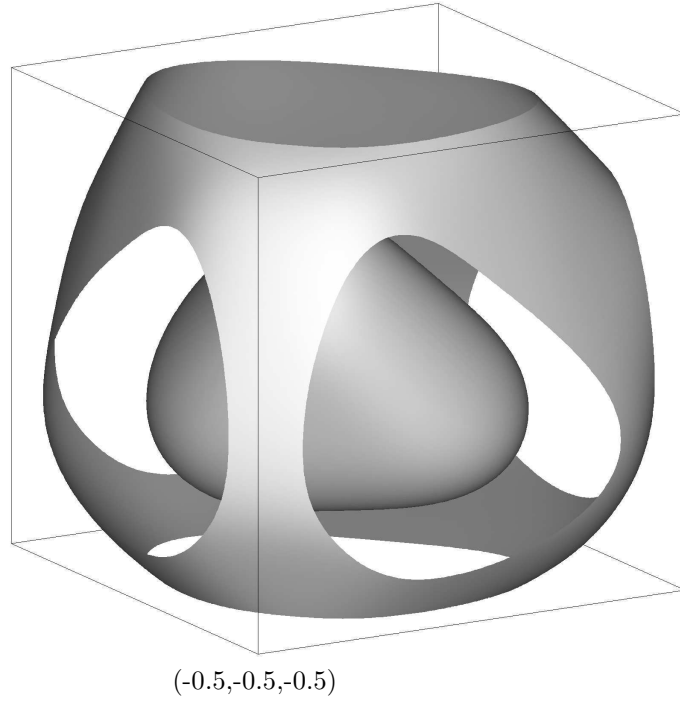
17

(-0.5,-0.5,-0.5)

Figure 2: Isosurfaces of the solution (7) to the test MAE (presented in Subsection 6.3) at the levels of a half and 1/8 of the maximum. The periodicity cell $T^3([0,1]^3)$ of u' is shown.
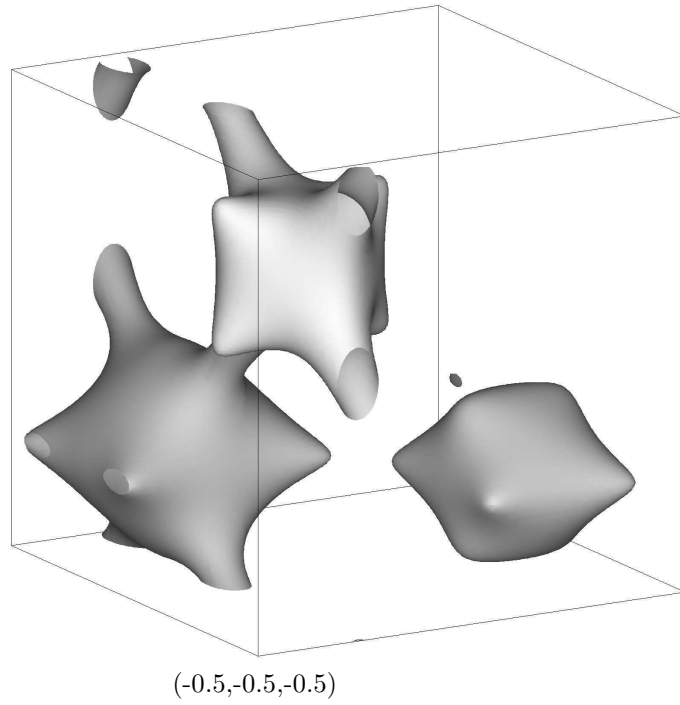

(-0.5,-0.5,-0.5)

Figure 3: Isosurfaces of $\nabla^2 u'$ for the solution to the test MAE at the level of 1/3 of the maximum. One periodicity cell $T^3([0,1]^3)$ is shown.
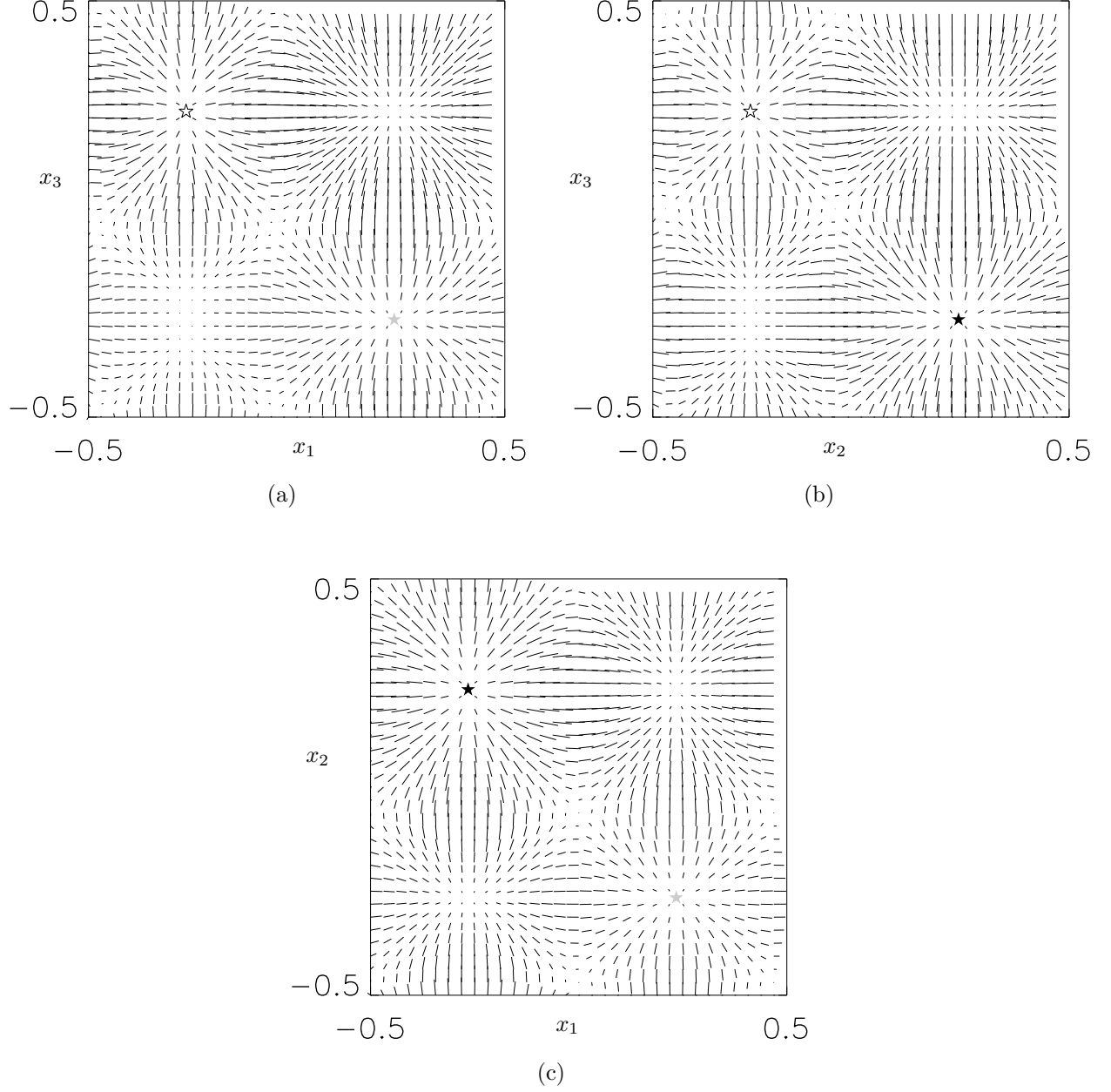
Figure 4: $\nabla u'$ for the solution to the test MAE (presented in Subsection 6.3) on cross sections of the periodicity cell $T^3([0,1]^3)$ that are parallel to coordinate planes and contain pairs of objects: $x_2 = -1/4$ (a), $x_1 = -1/4$ (b), $x_3 = -1/4$ (c). (Due to the symmetry of u' about each of the three planes, components of gradients normal to the planes are zero.) The labels x_i refer to the Cartesian coordinate axes, parallel to sides of the cross sections. Stars show locations of the three localised objects (36) on the cross sections. Gray-scaling reflects the masses of the objects (black, gray and white stars: the objects at $\mathbf{r}^{(g)}$, $g = 1, 2, 3$, respectively).
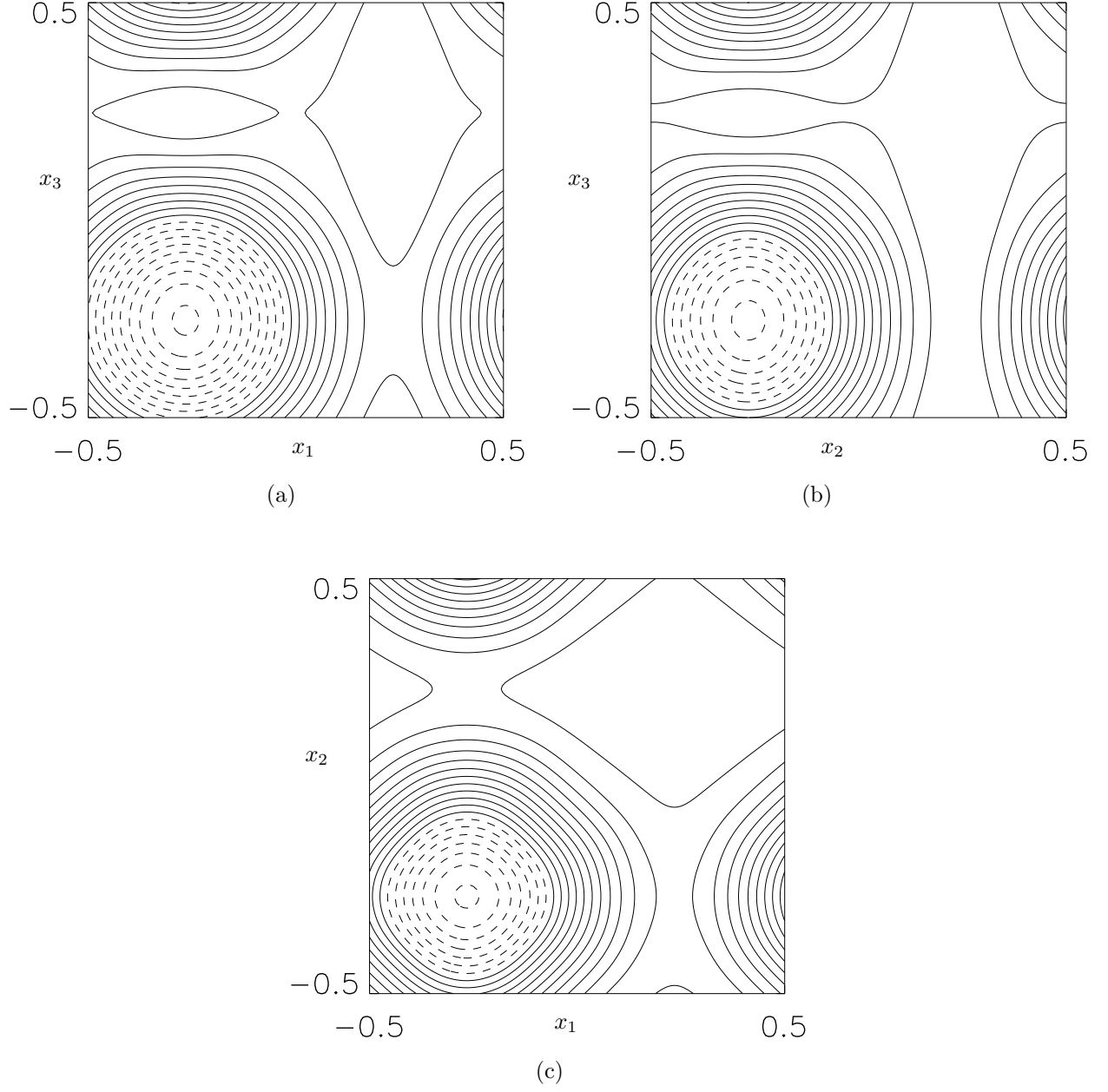
Figure 5: Isolines step 0.02 of normal components of $\nabla u'$ for the solution to the test MAE (dashed lines: negative values, solid lines: zero and positive values) on Cartesian coordinate planes $x_2 = 0$ (a), $x_1 = 0$ (b), $x_3 = 0$ (c). The labels x_i refer to the Cartesian coordinate axes, parallel to sides of the shown cross sections of the periodicity cell $T^3([0, 1]^3)$.

1/16 from this point. The Hessian of (7) computed for the spurious minimum solution is negative at this point, i.e. this solution (7) is not a convex function, as it has to be [13]. This has suggested to develop a modification of ACPDM, presented here, which we call AICDM (*Algorithm with Improvement of Convexity and Discrepancy Minimisation*).

The algorithm proceeds as an extension of the ACPDM operating with a single node $p = 1$; computation of a good initial approximation by extrapolation in p is unnecessary. AICDM involves an additional procedure: improvement of convexity of an approximate solution, which is performed once the condition $d^2(\eta_K) < \beta d^2(\eta_{K'})$ is satisfied. Here K is the number of the current iteration, K' is the number of the iteration, at which the last previous improvement of convexity was performed, and $\beta < 1$ is a constant factor (we have chosen $\beta = 0.01$ in the test computation reported in this subsection).

Improvement of convexity in AICDM could be performed by computation of a convex hull of the iterate, but we prefer to apply a numerically simpler procedure. At each grid point in the physical space, we compute the eigenvalues λ of the Hessian of $u' = \nabla^{-2}(\eta_K - \langle\eta_K\rangle)$ (they are all real, since the Hessian is a symmetric matrix). If they are all larger than -1, then the approximate solution given by (7) for the current iterate η_K is locally convex at the point under consideration (in this discussion we assume that, after normalisation, $c = 1$). Accordingly, if the minimum eigenvalue $\lambda_{\min}$ exceeds -1, no action is taken. Suppose now the contrary, i.e. $\lambda_{\min} < -1$. The minimum eigenvalue would become -1, if at this point each second derivative $u'_{x_i x_i}$ is increased by $-1 - \lambda_{\min}$, and hence the Laplacian $\eta_K = \nabla^2 u'$ is increased by $3(-1 - \lambda_{\min})$. Following this observation, at the points where $\lambda_{\min} < -1$ we increase η_K by $6(-1 - \lambda_{\min})$ (by choosing the factor 6 instead of 3 we are "overimproving" η_K). It is not guaranteed, of course, that $|\mathbf{r}|^2/2 + \nabla^{-2}(\eta'_K - \langle\eta'_K\rangle)$ is convex for the resultant η'_K, since this procedure changes all mixed derivatives at each point and does not guarantee that each second derivative $u'_{x_i x_i}$ at the points of local non-convexity is increased by the same amount, or that non-convexity does not appear at new grid points. Hence we proceed, repeating the procedure till each eigenvalue at each grid point becomes larger than $-1 - d(\eta_K)/2$. In our experience, only a small number of such iterations is necessary to enforce this condition. The quantities $6(-1 - \lambda_{\min})$ are small, being at most comparable with the global discrepancy $d(\eta_K)$; although the discrepancy $d(\eta'_K)$ for the "improved" approximation η'_K can exceed the discrepancy $d(\eta_K)$ for the original approximation η_K, it thus turns out that the increase (compared to $d(\eta_K)$) is usually modest.

An approximate solution to the test problem (formulated in Subsection 6.3), satisfying $d(\eta) = 10^{-10}$ and $d_\infty(\eta) = 0.41 \times 10^{-8}$, has been obtained by the AICDM. This has required $2\,653$ evaluations of the determinant of the Hessian. Table 1 illustrates the deterioration of convergence in this run, as better accuracy numerical solutions are successively found. Note that computational cost of each iteration in improvement of convexity of an approximate solution slightly exceeds that of computation of the determinant of the Hessian, and each iteration is included into evaluation counts (one evaluation per a convexity improving iteration) presented in Tables 1 and 2.

In this run, as in runs reported in Subsection 6.3, the Lebesgue space scalar product (33) has been assumed. We have also inspected convergence of AICDM employing scalar products

$$(\mathbf{u}, \mathbf{v}) = \int_{T^3} u(\mathbf{x})v(\mathbf{x})w(\mathbf{x})d\mathbf{x}$$

with weight functions

$$w(x; q) = \max\left(1, (\widehat{f}/\langle\widehat{f}\rangle)^q\right) \tag{38}$$

(see Table 2). For $q > 0$, more prominence is given to discrepancy in the regions, where the r.h.s. of the MAE, $\widehat{f}$, admits relatively high values; for $q < 0$, the opposite happens, i.e. discrepancy in the regions, where the values of $\widehat{f}$ are relatively low, is given more weight. Duration of the shortest run, for $q = -1/2$, with a sequential code on a 3.16 GHz Intel Core Duo processor is 25 min. 19 sec. The computations do not reveal any clear dependence of the efficiency of AICDM on the power q — in a large interval of q the variation of the number of iterations is just several per cent. The low sensitivity to the value of $q > 0$ is probably linked to the relative smallness of the region where $\widehat{f} > \langle\widehat{f}\rangle$.

6.5. An enhanced version of AICDM

The efficiency of AICDM can be further improved without employing new mathematical ideas. For instance, we have explored the possibility of gradual refinement of approximate solutions with increasing spatial resolution. The test problem, considered in the two previous subsections, was solved with the resolution of $(16M)^3$ Fourier harmonics, where the integer M is successively increased from 1 to 4. When seeking a solution with the resolution 64^3 harmonics, which satisfies the accuracy requirement $d(\eta) < 10^{-\kappa}$, the intermediate-stage computations with the resolution of $(16M)^3$ harmonics are terminated as soon as AICDM finds a solution to the accuracy

$$d(\eta) < 10^{-\min(\kappa, 2.5M)}.$$

We call this algorithm enhanced AICDM, or EAICDM. Durations of runs for integer $\kappa \geq 3$ are shown in Table 2 (we do not present total numbers of evaluations of the determinant of the Hessian in runs by EAICDM, since the times of their computations vary significantly with the resolution).

A further acceleration is likely to be achieved by avoiding dealiasing, but we did not explore this possibility.

7. Concluding remarks

We have presented new forms of the Monge–Ampère equation in R^N: the second-order divergence (4), Fourier integral (11) and convolution (21) forms. They have been derived under the assumption that the MAE (1) has a r.h.s. with a non-vanishing spatial mean, and hence (1) admits solutions (7). The first form gives an opportunity to relax the regularity requirements for weak solutions to the MAE, such that the integral in the l.h.s. of the identity (5) is well-defined. The third form is an integro-differential equation of the first order, which might be useful for development of an algorithm for computation of a solution (7) using transformations of spatial variables. This paper is mostly concerned with the Fourier integral form of the MAE used to prove existence of a small-amplitude solution of the form (7) for a weakly varying r.h.s. in (1) and to formulate an algorithm for computation of a solution (7) to (1) in an odd-dimensional space. In a test application to a three-dimensional MAE with the r.h.s. reminiscent of mass transportation problems considered in cosmology, we have demonstrated that a solution to the MAE with a smooth positive r.h.s. can be efficiently obtained by two versions of this algorithm, ACPDM (the algorithm with continuation in a parameter and discrepancy minimisation) and AICDM (the algorithm with improvement of convexity and discrepancy minimisation). While the latter is suitable for computation of a convex solution for an everywhere positive r.h.s. f, the former requires only $\langle f \rangle \neq 0$. However, modification of ACPDM for the case of a zero-mean space-periodic r.h.s. is straightforward.

There are some analogies between our method and the inexact Newton–Krylov solver with preconditioning, used in [21]. In our algorithms, solving (27) or (32), where the easily invertible polynomial part of the MAE is separated out, can be regarded as preconditioning; we believe that it is optimal. Indeed, its design is based on the algebraic nature of the MAE and primarily on the positivity and bounds of the kernels in the Fourier integral form, a very special — and so far not reported — property of the MAE. Furthermore, our algorithms are of the Krylov type. However, we do not rely on Newton iterations: instead of solving a succession of Newton problems, each with a complexity comparable to that of the MAE, we tackle the MAE directly.

In this paper, space periodicity and Fourier decompositions play an important rôle in two respects: First, we rely on the Fourier methods to find the optimal coefficients (28) in the equations (27) or (32). Note that space periodicity is not required for the derivations — we consider Fourier integrals, and not Fourier sums, and the algorithms do not require computation of these integrals. Second, our algorithms are realised in the spectral form, because we seek space-periodic solutions. In a periodicity domain the required computations of the inverse Laplacian, as well as numerical differentiation become trivial, when spectral methods are applied. Recall that Fourier methods have also the advantage of being more accurate: when the resolution is increased, Fourier series converge to a solution together with all the derivatives that the solution possesses, whilst finite differences do not approximate derivatives beyond their fixed order. We

would also use the spectral approach, for instance, to solve numerically Dirichlet or Neumann boundary value problems for u' (see (7)) in a rectangular box.

However, the geometry of the domain, where the solution is sought, may prohibit the use of spectral methods. What happens when the MAE must be solved for more complicated boundary conditions in regions whose geometry is not rectangular (parallelepiped-shaped)? Our algorithms do not inherently rely on spectral methods and will remain applicable. Since we are solving the MAE in terms of the Laplacian of the unknown function, we would have to invert the Laplacian with suitable regular boundary conditions. Finite differences can be applied in conjunction with our method to carry out this task and for computation of derivatives involved in the function F appearing in the r.h.s. of (32) and (32). Key equations for our algorithms, (27) and (32), are formulated in the physical space and can be used in conjunction with our method for a variety of boundary conditions. Their convergence properties remain of course to be investigated — for instance, whether the theoretically predicted values (28) will still be the optimal choice for the coefficients of the polynomial in the l.h.s. of (32).

An attractive feature of our algorithm is its simplicity: our Fortran-95 source code realising EAICDM is below 18 KB (672 lines long, not including the source for the Fast Fourier Transform).

Hereafter we list some open questions. Under which conditions does the generalised MAE

$$(1 - p) \sum_{m=0}^{N} a_m \eta^m + p \det \|\nabla^{-2} \eta_{x_i x_j} + \delta_{ij}\| = f/\langle f \rangle \tag{39}$$

with the coefficients (28), which is solved for a set of p by ACPDM, possess zero-mean space-periodic solutions for all $0 < p < 1$? Does its discretisation always have a solution, as long as the generalised MAE itself does? Does the solution of (39) depend analytically on the parameter p, and hence polynomial extrapolation for $p = 1$, that we employ, is mathematically sensible, or should another asymptotics near $p = 1$ be assumed? What is the optimal choice of the sequence of values of p to be used by ACPDM? Which scalar product is optimal for acceleration of convergence of stabilised iterations? Can Chebyshev techniques [32] be used to improve efficiency of the iterative processes?

Additional questions emerging outside the main topic of this paper — numerical methods for solution of the MAE — also cannot be avoided, since any information concerning the structure of solutions to Monge–Ampère problems of the cosmological type with a high-contrast r.h.s., formulated in Section 5, can be incorporated into specialised solvers in order to improve their performance (like it has proved possible to accelerate computations about 4 times just by taking into account in AICDM convexity of solutions to the MAE with a positive r.h.s.). The questions are: What is the asymptotics of solutions in the small parameter δ determining the width of the localised objects? Figures 4 and 5 show that the space is divided into regions of dominant influence of each object. Also, notable are almost axisymmetric structures in the Laplacian of the solution around the lines connecting centres of objects, seen on Figure 3. Can these geometric features be identified by performing the asymptotic analysis of the problem? How does the contrast number measure numerical complexity of the MAE and, in particular, the condition number of the linearisation near the solution (for a given spatial discretisation)?

References

[1] S. Haker, A. Tannenbaum, R. Kikinis. Mass preserving mappings and image registration. Proc. of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention. Lect. Notes in Comput. Sci., vol. 2208. Springer-Verlag, London (2001) 120–127.

[2] S. Haker, L. Zhu, A. Tannenbaum, S. Angement. Optimal mass transport for registration and warping. Int. J. Comput. Vision, 60 (2004) 225-240.

[3] T. Hurtut, Y. Gousseau, F. Schmitt. Adaptive image retrieval based on the spatial organization of colors. Computer vision and image understanding, 112 (2008) 101–113.

[4] U. Frisch, S. Matarrese, R. Mohayaee, A. Sobolevski. A reconstruction of the initial conditions of the Universe by optimal mass transportation. Nature, 417 (2002) 260–262 (arXiv:astro-ph/0109483).

[5] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, A. Sobolevskii. Reconstruction of the early Universe as a convex optimization problem. Mon. Not. R. Astron. Soc., 346 (2003) 501–524 (arXiv:astro-ph/0304214).

[6] R. Mohayaee, A. Sobolevskii, The Monge–Ampère–Kantorovich approach to reconstruction in cosmology. Physica D, 237 (2008) 2145–2150 (arXiv:0712.2561).

[7] D. Gilbarg, N.S. Trudinger. Elliptic partial differential equations of second order. Springer-Verlag, Berlin, 1983.

[8] I.J. Bakelman. Convex analysis and nonlinear geometric elliptic equations. Springer-Verlag, 1994.

[9] L.A. Caffarelli, X. Cabré. Fully nonlinear elliptic equations. American Mathematical Society colloquium publications, vol. 43. Amer. Math. Soc., Providence, Rhode Island, 1995.

[10] V.I. Oliker, L.D. Prussner. On the numerical solution of the equation $\dfrac{\partial^2 z}{\partial x^2} \dfrac{\partial^2 z}{\partial y^2} - \left(\dfrac{\partial^2 z}{\partial x \partial y} \right)^2 = f$ and its discretizations, I. Numerische Mathematik, 54 (1988) 271–293.

[11] D. Michaelis, S. Kudaev, R. Steinkopf, A. Gebhardt, P. Schreiber, A. Bräuer. Incoherent beam shaping with freeform mirror. Nonimaging optics and efficient illumination systems V. Eds. R. Winston, R.J. Koshel. Proc. of SPIE, vol. 7059 (2008) 705905.

[12] T. Glimm, V. Oliker. Optical design of single reflector systems and the Monge–Kantorovich mass transfer problem. J. Math. Sci. 117 (2003) 4096–4108.

[13] A.V. Pogorelov. The Minkowski multidimensional problem. Halsted Press, Washington D.C., 1978. (Transl. from Russian: A.V. Pogorelov, The Minkowski multidimensional problem. Nauka, Moscow, 1975.)

[14] J.-D. Benamou, Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numerische Mathematik, 84 (2000) 375–393.

[15] E.J. Dean, R. Glowinski. Numerical solution of the two-dimensional elliptic Monge–Ampère equation with Dirichlet boundary conditions: an augmented Lagrangian approach. C. R. Acad. Sci. Paris, Ser. I, 336 (2003) 779–784.

[16] E.J. Dean, R. Glowinski. Numerical solution of the two-dimensional elliptic Monge–Ampère equation with Dirichlet boundary conditions: a least-squares approach. C. R. Acad. Sci. Paris, Ser. I, 339 (2004) 887–892.

[17] X. Feng, M. Neilan. Galerkin methods for the fully nonlinear Monge–Ampère equation (arXiv:0712.1240).

[18] X. Feng, M. Neilan. Mixed finite element methods for the fully nonlinear Monge–Ampère equation based on the vanishing moment method (arXiv:0712.1241).

[19] G. Loeper, F. Rapetti. Numerical solution of the Monge–Ampère equation by a Newton's algorithm. C. R. Acad. Sci. Paris, Ser. I, 340 (2005) 319–324.

[20] J.-D. Benamou, B.D. Froese, A.M. Oberman. Two numerical methods for the elliptic Monge–Ampère equation. Preprint, 2009 [www.divbyzero.ca/froese/w/images/4/40/MA.pdf].

[21] G.L. Delzanno, L. Chacón, J.M. Finn, Y. Chung, G. Lapenta. An optimal robust equidistribution method for two-dimensional grid adaptation based on Monge–Kantorovich optimization. J. Comput. Physics, 227 (2008) 9841–9864.

[22] J.M. Finn, G.L. Delzanno, L. Chacon. Grid generation and adaptation by Monge–Kantorovich optimization in two and three dimensions. Proceedings of the 17th International Meshing Roundtable (2008) 551–568.

[23] C.E. Gutiérrez. The Monge-Ampère Equation. Progress in nonlinear differential equations and their applications, vol. 44. Birkhäuser, Boston, 2001.

[24] A.-M. Ampère. Mémoire concernant ... l'intégration des équations aux différentielles partielles du premier et du second ordre. Journal de L'École Royale Polytechnique 11 (1820) 1–188.

[25] E.J. Dean, R. Glowinski. Numerical methods for fully nonlinear elliptic equations of the Monge–Ampère type. Comput. Methods Appl. Mech. Engrg. 195 (2006) 1344–1386.

[26] A.I. Kostrikin. Introduction to algebra. Nauka, Moscow, 1977 (in Russian).

[27] P.J.E. Peebles. Tracing galaxy orbits back in time. Astrophys. J., 344 (1989) L53–L56.

[28] Ya.B. Zel'dovich. Gravitational instability: an approximate theory for large density perturbations. Astron. & Astrophys. 5 (1970), 84–89.

[29] F. Moutarde, J.-M. Alimi, F.R. Bouchet, R. Pellat, A. Ramani. Precollapse scale invariance in gravitational instability. Astrophys. J. 382 (1991) 377–381.

[30] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteur. C.R. Acad. Sci. Paris, Ser. I, 305 (1987) 805–808.

[31] O. Axelsson. Iterative solution methods. Cambridge Univ. Press, 1996.

[32] O.M. Podvigina, V.A. Zheligovsky. An optimized iterative method for numerical solution of large systems of equations based on the extremal property of zeroes of Chebyshev polynomials. J. Scientific Computing, 12 (1997) 433–464.